EXHIBIT F

TO DECLARATION OF S. MERRILL WEISS IN SUPPORT OF PLAINTIFF ACACIA MEDIA TECHNOLOGIES CORPORATION'S MEMORANDUM OF POINTS AND AUTHORITIES IN OPPOSITION TO ROUND 3 DEFENDANTS' MOTION FOR SUMMARY JUDGMENT OF INVALIDITY UNDER 35 U.S.C. § 112 OF THE '992, '863, AND '702 PATENTS; AND SATELLITE DEFENDANTS' MOTION FOR SUMMARY JUDGMENT OF INVALIDITY OF THE '992, '863, AND '720 PATENTS

Hovsepian v. Apple, Inc.

Doc. 325 Att. 5

Dockets.Justia.com

A Method for the Construction of Minimum-Redundancy Codes*

DAVID A. HUFFMAN†, ASSOCIATE, IRE

Summary—**An optimum method of coding an ensemble of messages consisting of a finite number of members is developed. A minimum-redundancy code is one constructed in such a way that the average number of coding digits per message is minimized.**

Introduction

One important method of transmitting messages is to transmit in their place sequences of symbols. If there are more messages which might be sent than there are kinds of symbols available, then some of the messages must use more than one symbol. If it is assumed that each symbol requires the same time for transmission, then the time for transmission (length) of a message is directly proportional to the number of symbols associated with it. In this paper, the symbol or sequence of symbols associated with a given message will be called the "message code." The entire number of messages which might be transmitted will be called the "message ensemble." The mutual agreement between the transmitter and the receiver about the meaning of the code for each message of the ensemble will be called the "ensemble code."

Probably the most familiar ensemble code was stated in the phrase "one if by land and two if by sea." In this case, the message ensemble consisted of the two individual messages "by land" and "by sea", and the message codes were "one" and "two."

In order to formalize the requirements of an ensemble code, the coding symbols will be represented by numbers. Thus, if there are D different types of symbols to be used in coding, they will be represented by the digits $0, 1, 2, \cdots, (D-1)$. For example, a ternary code will be constructed using the three digits 0, 1, and 2 as coding symbols.

The number of messages in the ensemble will be called N. Let $P(i)$ be the probability of the ith message. Then

$$\sum_{i=1}^{N} P(i) = 1. \qquad (1)$$

The length of a message, $L(i)$, is the number of coding digits assigned to it. Therefore, the average message length is

$$L_{av} = \sum_{i=1}^{N} P(i)L(i). \qquad (2)$$

The term "redundancy" has been defined by Shannon[1] as a property of codes. A "minimum-redundancy code" will be defined here as an ensemble code which, for a message ensemble consisting of a finite number of members, N, and for a given number of coding digits, D, yields the lowest possible average message length. In order to avoid the use of the lengthy term "minimum-redundancy," this term will be replaced here by "optimum." It will be understood then that, in this paper, "optimum code" means "minimum-redundancy code."

The following basic restrictions will be imposed on an ensemble code:

(a) No two messages will consist of identical arrangements of coding digits.

(b) The message codes will be constructed in such a way that no additional indication is necessary to specify where a message code begins and ends once the starting point of a sequence of messages is known.

Restriction (b) necessitates that no message be coded in such a way that its code appears, digit for digit, as the first part of any message code of greater length. Thus, 01, 102, 111, and 202 are valid message codes for an ensemble of four members. For instance, a sequence of these messages 1111022020101111102 can be broken up into the individual messages 111-102 202-01-01-111-102. All the receiver need know is the ensemble code. However, if the ensemble has individual message codes including 11, 111, 102, and 02, then when a message sequence starts with the digits 11, it is not immediately certain whether the message 11 has been received or whether it is only the first two digits of the message 111. Moreover, even if the sequence turns out to be 11102, it is still not certain whether 111-02 or 11-102 was transmitted. In this example, change of one of the two message codes 111 or 11 is indicated.

C. E. Shannon[1] and R. M. Fano[2] have developed ensemble coding procedures for the purpose of proving that the average number of binary digits required per message approaches from above the average amount of information per message. Their coding procedures are not optimum, but approach the optimum behavior when N approaches infinity. Some work has been done by Kraft[3] toward deriving a coding method which gives an average code length as close as possible to the ideal when the ensemble contains a finite number of members. However, up to the present time, no definite procedure has been suggested for the construction of such a code

* Decimal classification: R531.1. Original manuscript received by the Institute, December 6, 1951.

† Massachusetts Institute of Technology, Cambridge, Mass.

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. Jour.*, vol. 27, pp. 398–403; July, 1948.

[2] R. M. Fano, "The Transmission of Information," Technical Report No. 65, Research Laboratory of Electronics, M.I.T., Cambridge, Mass.; 1949.

[3] L. G. Kraft, "A Device for Quantizing, Grouping, and Coding Amplitude-modulated Pulses," Electrical Engineering Thesis, M.I.T., Cambridge, Mass.; 1949.

Exhibit F Page 1066

to the knowledge of the author. It is the purpose of this paper to derive such a procedure.

Derived Coding Requirements

For an optimum code, the length of a given message code can never be less than the length of a more probable message code. If this requirement were not met, then a reduction in average message length could be obtained by interchanging the codes for the two messages in question in such a way that the shorter code becomes associated with the more probable message. Also, if there are several messages with the same probability, then it is possible that the codes for these messages may differ in length. However, the codes for these messages may be interchanged in any way without affecting the average code length for the message ensemble. Therefore, it may be assumed that the messages in the ensemble have been ordered in a fashion such that

$$P(1) \geqq P(2) \geqq \cdots \geqq P(N-1) \geqq P(N) \qquad (3)$$

and that, in addition, for an optimum code, the condition

$$L(1) \leqq L(2) \leqq \cdots \leqq L(N-1) \leqq L(N) \qquad (4)$$

holds. This requirement is assumed to be satisfied throughout the following discussion.

It might be imagined that an ensemble code could assign q more digits to the Nth message than to the $(N-1)$st message. However, the first $L(N-1)$ digits of the Nth message must not be used as the code for any other message. Thus the additional q digits would serve no useful purpose and would unnecessarily increase L_{av}. Therefore, for an optimum code it is necessary that $L(N)$ be equal to $L(N-1)$.

The kth prefix of a message code will be defined as the first k digits of that message code. Basic restriction (b) could then be restated as: No message shall be coded in such a way that its code is a prefix of any other message, or that any of its prefixes are used elsewhere as a message code.

Imagine an optimum code in which no two of the messages coded with length $L(N)$ have identical prefixes of order $L(N)-1$. Since an optimum code has been assumed, then none of these messages of length $L(N)$ can have codes or prefixes of any order which correspond to other codes. It would then be possible to drop the last digit of all of this group of messages and thereby reduce the value of L_{av}. Therefore, in an optimum code, it is necessary that at least two (and no more than D) of the codes with length $L(N)$ have identical prefixes of order $L(N)-1$.

One additional requirement can be made for an optimum code. Assume that there exists a combination of the D different types of coding digits which is less than $L(N)$ digits in length and which is not used as a message code or which is not a prefix of a message code. Then this combination of digits could be used to replace the code for the Nth message with a consequent reduction of L_{av}. Therefore, all possible sequences of $L(N)-1$ digits must be used either as message codes, or must have one of their prefixes used as message codes.

The derived restrictions for an optimum code are summarized in condensed form below and considered in addition to restrictions (a) and (b) given in the first part of this paper:

(c) $$L(1) \leqq L(2) \leqq \cdots \leqq L(N-1) = L(N). \qquad (5)$$

(d) At least two and not more than D of the messages with code length $L(N)$ have codes which are alike except for their final digits.

(e) Each possible sequence of $L(N)-1$ digits must be used either as a message code or must have one of its prefixes used as a message code.

Optimum Binary Code

For ease of development of the optimum coding procedure, let us now restrict ourselves to the problem of binary coding. Later this procedure will be extended to the general case of D digits.

Restriction (c) makes it necessary that the two least probable messages have codes of equal length. Restriction (d) places the requirement that, for D equal to two, there be only two of the messages with coded length $L(N)$ which are identical except for their last digits. The final digits of these two codes will be one of the two binary digits, 0 and 1. It will be necessary to assign these two message codes to the Nth and the $(N-1)$st messages since at this point it is not known whether or not other codes of length $L(N)$ exist. Once this has been done, these two messages are equivalent to a single composite message. Its code (as yet undetermined) will be the common prefixes of order $L(N)-1$ of these two messages. Its probability will be the sum of the probabilities of the two messages from which it was created. The ensemble containing this composite message in the place of its two component messages will be called the first auxiliary message ensemble.

This newly created ensemble contains one less message than the original. Its members should be rearranged if necessary so that the messages are again ordered according to their probabilities. It may be considered exactly as the original ensemble was. The codes for each of the two least probable messages in this new ensemble are required to be identical except in their final digits; 0 and 1 are assigned as these digits, one for each of the two messages. Each new auxiliary ensemble contains one less message than the preceding ensemble. Each auxiliary ensemble represents the original ensemble with full use made of the accumulated necessary coding requirements.

The procedure is applied again and again until the number of members in the most recently formed auxiliary message ensemble is reduced to two. One of each of the binary digits is assigned to each of these two composite messages. These messages are then combined to form a single composite message with probability unity, and the coding is complete.

Exhibit F Page 167

TABLE I
OPTIMUM BINARY CODING PROCEDURE

Message Probabilities												
Original Message Ensemble	*Auxiliary Message Ensembles*											
	1	2	3	4	5	6	7	8	9	10	11	12
												→1.00
											→0.60	
										→0.40	0.40	
									→0.36	0.36		
								→0.24	0.24	0.24		
0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20			
							→0.20	0.20	0.20			
0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18				
						→0.18	0.18	0.18				
					→0.14	0.14	0.14					
0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10					
0.10	0.10	0.10	0.10	0.10	0.10	0.10						
0.10	0.10	0.10	0.10	0.10	0.10	0.10						
				→0.10	0.10							
		→0.08	0.08	0.08	0.08							
			→0.08	0.08								
0.06	0.06	0.06	0.06	0.06								
0.05	0.06	0.06	0.06									
0.04	0.04	0.04	0.04									
*0.04	0.04	0.04										
0.04	0.04	0.04										
0.04	0.04											
	→0.04											
0.03												
0.01												

Now let us examine Table I. The left-hand column contains the ordered message probabilities of the ensemble to be coded. N is equal to 13. Since each combination of two messages (indicated by a bracket) is accompanied by the assigning of a new digit to each, then the total number of digits which should be assigned to each original message is the same as the number of combinations indicated for that message. For example, the message marked *, or a composite of which it is a part, is combined with others five times, and therefore should be assigned a code length of five digits.

When there is no alternative in choosing the two least probable messages, then it is clear that the requirements, established as necessary, are also sufficient for deriving an optimum code. There may arise situations in which a choice may be made between two or more groupings of least likely messages. Such a case arises, for example, in the fourth auxiliary ensemble of Table I. Either of the messages of probability 0.08 could have been combined with that of probability 0.06. However, it is possible to rearrange codes in any manner among equally likely messages without affecting the average code length, and so a choice of either of the alternatives could have been made. Therefore, the procedure given is always sufficient to establish an optimum binary code.

The lengths of all the encoded messages derived from Table I are given in Table II.

Having now determined proper lengths of code for each message, the problem of specifying the actual digits remains. Many alternatives exist. Since the combining of messages into their composites is similar to the successive confluences of trickles, rivulets, brooks, and creeks into a final large river, the procedure thus far described might be considered analogous to the placing of signs by a water-borne insect at each of these junctions as he journeys downstream. It should be remembered that the code which we desire is that one which the insect must remember in order to work his way back upstream. Since the placing of the signs need not follow the same rule, such as "zero-right-returning," at each junction, it can be seen that there are at least 2^{12} different ways of assigning code digits for our example.

TABLE II
RESULTS OF OPTIMUM BINARY CODING PROCEDURE

i	$P(i)$	$L(i)$	$P(i)L(i)$	Code
1	0.20	2	0.40	10
2	0.18	3	0.54	000
3	0.10	3	0.30	011
4	0.10	3	0.30	110
5	0.10	3	0.30	111
6	0.06	4	0.24	0101
7	0.06	5	0.30	00100
8	0.04	5	0.20	00101
9	0.04	5	0.20	01000
10	0.04	5	0.20	01001
11	0.04	5	0.20	00110
12	0.03	6	0.18	001110
13	0.01	6	0.06	001111
			$L_{av}=3.42$	

The code in Table II was obtained by using the digit 0 for the upper message and the digit 1 for the lower message of any bracket. It is important to note in Table I that coding restriction (e) is automatically met as long as two messages (and not one) are placed in each bracket.

Exhibit F Page 168

Generalization of the Method

Optimum coding of an ensemble of messages using three or more types of digits is similar to the binary coding procedure. A table of auxiliary message ensembles similar to Table I will be used. Brackets indicating messages combined to form composite messages will be used in the same way as was done in Table I. However, in order to satisfy restriction (e), it will be required that all these brackets, with the possible exception of one combining the least probable messages of the original ensemble, always combine a number of messages equal to D.

It will be noted that the terminating auxiliary ensemble always has one unity probability message. Each preceding ensemble is increased in number by $D-1$ until the first auxiliary ensemble is reached. Therefore, if N_1 is the number of messages in the first auxiliary ensemble, then $(N_1-1)/(D-1)$ must be an integer. However $N_1=N-n_0+1$, where n_0 is the number of the least probable messages combined in a bracket in the original ensemble. Therefore, n_0 (which, of course, is at least two and no more than D) must be of such a value that $(N-n_0)/(D-1)$ is an integer.

In Table III an example is considered using an ensemble of eight messages which is to be coded with four digits; n_0 is found to be 2. The code listed in the table is obtained by assigning the four digits 0, 1, 2, and 3, in order, to each of the brackets.

TABLE III

Optimum Coding Procedure for $D=4$

Message Probabilities					
Original Message Ensemble	*Auxiliary Ensembles*			$L(i)$	*Code*
			→1.00		
		→0.40			
0.22	0.22	0.22		1	1
0.20	0.20	0.20		1	2
0.18	0.18	0.18		1	3
0.15	0.15			2	00
0.10	0.10			2	01
0.08	0.08			2	02
	→0.07				
0.05				3	030
0.02				3	031

Acknowledgments

The author is indebted to Dr. W. K. Linvill and Dr. R. M. Fano, both of the Massachusetts Institute of Technology, for their helpful criticism of this paper.

Exhibit F Page 169